IATSS Research 39 (2015) 1–8

Contents lists available at ScienceDirect

IATSS Research

The challenges of transportation/traffic statistics in Japan and directions for the future

Shigeru Kawasaki

Department of Economics, Nihon University, Tokyo, Japan

ARTICLE INFO

Available online 4 July 2015

Keywords:
Official statistics
Transportation research
Microdata
Big data
Privacy

ABSTRACT

In order to respond to new challenges in transportation and traffic problems, it is essential to enhance statistics in this field that provides the basis for policy researches. Many of the statistics in this field in Japan consist of "official statistics" created by the government. This paper gives a review of the current status of transportation and traffic statistics (hereinafter called "transportation statistics" in short) in Japan. Furthermore, the paper discusses challenges in such statistics in the new environment and the direction that statistics that should take in the future. For Japan's transportation statistics to play vital roles in more sophisticated analyses, it is necessary to improve the environment that facilitates the use of microdata for analysis. It is also necessary to establish an environment where big data can be more easily used for compilation of official statistics and performing policy researches. To achieve this end, close cooperation among the government, academia, and businesses will be essential.

Contents

1. Introduction

Recent economic development and globalization have sparked increased demand for the movement and transportation of people and goods, creating the demand for safe and secure methods of transportation and services that meet users' needs for better quality and higher efficiency. Under such circumstances, the public and the private sectors

E-mail address: kawasaki.shigeru@nihon-u.ac.jp.

Peer review under responsibility of International Association of Traffic and Safety Sciences.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.iatssr.2015.06.002

and academic institutions have to research the actual conditions of transportation and traffic to find appropriate solutions for addressing new challenges. Reliable statistical data on transportation and traffic play essential roles in supporting such research. In Japan, numerous kinds of government-provided transportation statistics provide a foundation for various studies and analysis. Due to the rapid changes in society and the economy, however, the issues surrounding transportation and traffic complicate matters further; demands for statistical data have become more sophisticated and diverse, and it has become increasingly vital for scientific researches to get access to a broader range of statistics of higher quality.

Based on the above perspective, this paper explains the current transportation statistics in Japan and discusses approaches to creating data in the future. Official statistics created by the government play an important role in transportation statistics that researchers use for their analysis. The Statistics Act provides the basic framework of the official statistics of Japan, which comprises the transportation statistics [1].

In the following, I shall first explain the Statistics Act and the official statistical system. Next, I shall give a brief overview of Japanese transportation statistics. After laying out the main features and challenges of transportation statistics, I shall discuss possible directions for future improvements. More specifically, for existing official statistics, it is important to facilitate the use of microdata to make the most of their values. For developing new statistics, it is necessary to use big data as the data sources.

This paper aims to provide a basic reference for researchers looking to utilize the transportation statistics of Japan and makes proposals for further developments of Japanese transportation statistics. Although this discussion focuses on Japan, many of the issues that transportation statistics face are common worldwide; thus, I hope that this paper will contribute to the international debate on the improvement of transportation statistics.

2. The official statistical system and the Statistics Act

2.1. The official statistical system

The Japanese government creates various statistics about the Japanese society and economy, including transportation. In the official statistical system of Japan, the Statistics Bureau of the Ministry of Internal Affairs and Communications (MIC) serves as the core statistical agency, creating and publishing fundamental statistics, while individual ministries create and publish specialized statistics in their respective fields of jurisdiction. This kind of system for creating these statistics is called "decentralized statistics system."[1] Fig. 1 shows the primary statistical agencies in Japan's official statistics system (Fig. 1 here).

Under this system, the Ministry of Land, Infrastructure, Transport and Tourism (MLITT) is the main organization for creation and publication of official statistics on the actual conditions of transportation and traffic services. In addition, several other government agencies create relevant statistical data to facilitate research and analysis of transportation and traffic.

The statistical agencies responsible for official statistics publish their output on their websites as well as via printed reports. However, it is a difficult task for anyone to search all the websites, and to gather necessary statistics across sectors from different ministries. To facilitate cross-sectoral searches, the Statistics Bureau of the MIC works with other ministries to provide a one-stop service at the Statistical Portal Site of the Japanese Government (e-Stat)[2] as for the entire official statistical system [2]. With the e-Stat system, users can take advantage of cross-sectoral access to statistics, for example, through the "Regional Statistics Database," to perform crossover searches of various domains such as transportation statistics, demographic, and economic statistics and make a single comprehensive data set. Under a decentralized statistical system, there are measures that enable the use of unified statistical data across the government.

2.2. The Statistics Act

To ensure the reliability and utility of official statistics, it is essential to create and publish statistics under a government-wide uniform policy. For this purpose, the Statistics Act stipulates a common framework for all government statistics, including those on transportation. The first Statistics Act, enacted in 1947, set a clear policy of openness and truthfulness in official statistics with a view to prevent recurrence of the government's information control that took place during the Second World War. Since then, the Statistics Act provided the core framework for official statistics in Japan until the revision in 2007. As the time went by, however, some of its provisions fail to meet the changes in technology and user environment, and the Statistics Act was completely revised in 2007.

The section below summarizes the key features of the current Statistics Act.

The first feature is the basic philosophy on statistics. The Statistics Act defines the essential role of statistics as "critical information for the citizens in their rational decision making" (Article 1). The Act aims to contribute to the improvement of the economy and the living standards of the country's citizens by developing statistics systematically and efficiently. To materialize the law's goals, Article 3 stipulates four principles: (i) the systematic development of statistics, (ii) the assurance of neutrality and reliability, (iii) the provision of easy accessibility of statistics for citizens, and (iv) the protection of confidential matters concerning the populations surveyed. These principles are consistent with the "Fundamental Principles of Official Statistics," which the United Nations adopted in 1994 as a common international norm for official statistics [3]. The basic principles of the Statistics Act apply to all official statistics, and each article of the Act stems from these principles.

The second characteristic is the protection of confidentiality of individual data obtained by statistical surveys. If respondents have fears about confidentiality of the data they provide, they will not answer survey questions truthfully, and reliability of statistics thus obtained will be lost. For this reason, the Statistics Act stipulates strict protection of confidentiality and prohibits the use of individual data for non-statistical purposes.

Third, the Act stipulates that the Minister of Internal Affairs and Communications is to designate "Fundamental Statistical Surveys" for creating particularly important statistics within the system of statistics. Individuals and companies are required to respond to the survey, if they are asked. Through the Fundamental Statistical Surveys, the Act aims to obtain statistics with high accuracy. This is a continuation from "Designated Statistical Surveys," which the former Statistics Act defined. Fundamental Statistical Surveys comprise about 50 statistical surveys, such as "Population Census," "Economic Census," and, in the field of transportation, "Motor Vehicle Transportation Statistics." Statistical surveys other than the Fundamental Statistical Surveys are categorized as "ordinary statistical surveys" which are not mandatory for respondents.

The fourth feature is the formulation of the master plan for statistical development and the establishment of the Statistical Commission as an advisory body. The government determines the master plan for government-wide statistical development as the cabinet decision about every five years. The Statistics Commission is to give advice to the Prime Minister and the Minister of Internal Affairs and Communications on such occasions as designating new Fundamental Statistical

[1] Examples of the decentralized statistical systems can be found in the United States and the United Kingdom. On the other hand, the case where a single central statistical agency creates most of the statistics is called "centralized statistical system"; Canada, Australia and the Netherlands, for example, adopt this type of system. The distinction between "centralized" and "decentralized" depends on the degree of concentration, and the borderline is not always clear.

[2] The description of the e-Stat in this paper is based on the Japanese website. The English website has a limited functionality, and not all the functions described here are available. (Portal Site of Official Statistic of Japan (English version): http://www.e-stat.go.jp/SG1/estat/eStatTopPortalE.do).

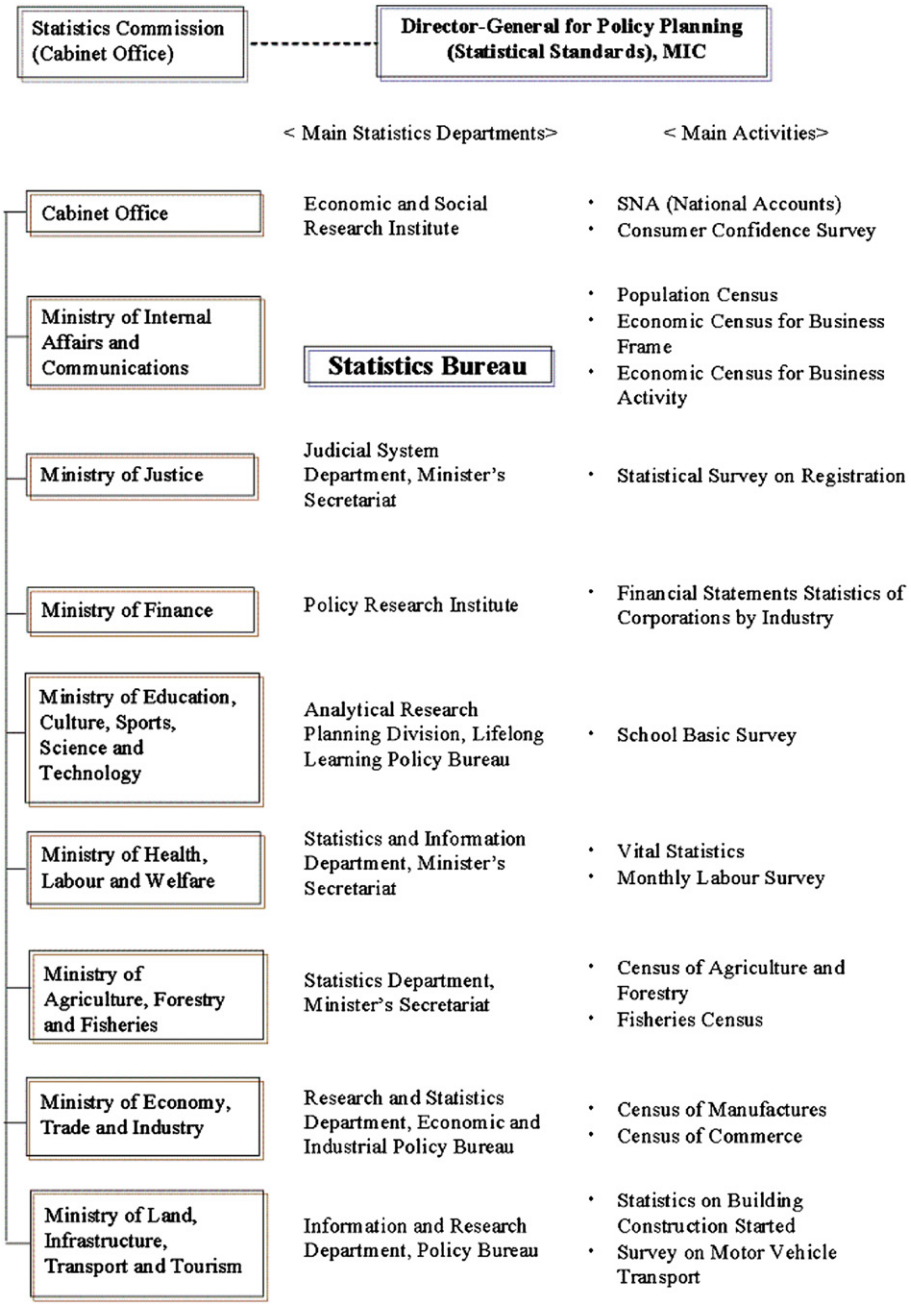

Fig. 1. Organizations in the official Statistical System of Japan (Source) Ministry of Internal Affairs and Communications, Japan http://www.soumu.go.jp/english/dgpp_ss/seido/1.htm.

Surveys, revising the Fundamental Statistical Surveys, and formulating the master plan. These procedures represent a mechanism for reflecting the opinions of a professional third party in the improvement of statistics.

Fifth, the law contains a stipulation on the use of microdata. The Statistics Act, while strictly protecting confidentiality of individual data in statistical surveys, allows the statistical offices to provide anonymized sample microdata files to researchers. In addition, the act also includes a provision for a fee-based custom-made tabulation service. In this way, the law opens a way to provide special data sets and services to meet the demands from researches.

The above provisions of the Statistics Act apply all fields of official statistics and statistical surveys, including transportation statistics.

2.3. Scope of statistics covered by the Statistics Law

The basic philosophy of the Statistics Act applies to all official statistics, but the direct implications of the Act depend on the type of the statistics in question. Official statistics normally fall into three general groups according to the source of information: (i) survey statistics, (ii) administrative statistics, and (iii) processed statistics. Survey statistics represent a collection of data gathered from households and corporations and/or establishments through the censuses and sample surveys such as the "Population Census" and "Household Income and Expenditure Survey." Administrative statistics cover information collected by the administrative agencies and made into statistics such as "International Trade Statistics" and "Crime Statistics." Processed statistics are

calculated by combining, and processing existing statistics to make estimations, such as the "National Accounts Statistics" and "Life Tables."

Of these three types, the Statistics Act has the most direct effect on survey statistics. Fundamental Statistical Surveys, which belongs to the "survey statistics" category, requires respondents to give answers to the surveys. When changes in the survey method are proposed, the Statistical Commission is asked by the Minister of Internal Affairs and Communications to express their opinions from the viewpoints such as improving the accuracy and the usability of the statistics and keeping the response burdens to a reasonable level. In contrast, ordinary statistical surveys do not give any obligation to respondents or operate under the strict procedure provisions that Fundamental Statistical Surveys call for, and the Statistical Commission is not asked to provide opinions.

Administrative statistics and processed statistics do not ask for responses from the public in creating the statistics, meaning that, unlike survey statistics, they do not provide explicit provisions relating to data collection.

3. Major transportation statistics

In official statistics, international statistical standards, such as the Standard Industrial Classification and the System of National Accounts, are deliberated and adopted by the United Nations Statistical Commission (UNSC), and most of the countries in the world observe such standards. In transportation statistics, there has not been an international statistical standard formally agreed at the UNSC, and each country creates statistics to meet the needs of its respective society and economy.

In major developed countries, governments provide transportation statistics in print or online. In Japan, the MLITT provides "Transportation-Related Statistics Data" on the Internet. In addition, the Statistics Bureau of the MIC provides statistics on transportation in the *Japan Statistical Yearbook* in "Chapter 12: Transport and Tourism" and "Chapter 26: Environment, Disasters and Accidents". In these materials, transportation statistics can be classified into four groups: (i) statistics on the supply and demand of transportation services, (ii) statistics on traffic accidents and safety, (iii) statistics on energy and the environment related to transportation services, and (iv) statistics on the movement of people.

Applying this scheme, the following section provides an overview of how Japanese transportation statistics have been provided.

3.1. Statistics on the supply and demand of transportation

The most fundamental of all transportation statistics are those on the supply and demand of transportation, most of which have been created by the MLITT. The ministry's transportation statistics are available on the "statistics" page of its website (Japanese version only). Table 1 lists the main statistics available on the site.

In broad terms, these statistics comprise data on transportation services, transportation infrastructure, and consumers of transportation.

Statistics on transportation services are generally created by collecting information from suppliers of such services. Many of the statistics in this field come from ordinary statistical surveys, and only a small portion comes from Fundamental Statistical Surveys. This is because there are often legal regulations and reporting systems that govern businesses in this sector. In statistical surveys conducted by the MLITT, which is the regulating body for the transportation sector, businesses tend to feel obliged to respond even if no legal obligation is stipulated. Because of such circumstances, transportation statistics include many ordinary statistical surveys, administrative statistics based on laws other than the Statistics Act, and processed statistics and other data created in administrative processes.

Statistics concerning the transportation infrastructure of roads, ports, railways, and other similar facilities rely on administrative statistics rather than survey statistics because there are often established administrative systems that require public agencies and relatively large-scale railway operators to report to the government.

Table 1
The main transportation statistics of the MLITT.

Policy area	Main statistics
General/other	Metropolitan Transportation Census (ordinary)
	Inter-Regional Travel Survey in Japan (ordinary)
	Net Freight Flow Census in Japan (logistics census) (ordinary)
	Inter-Regional Passenger Mobility Survey (processed)
	Report on Cargo Flow in Japan (processed)
City	Nationwide Person Trip Survey (ordinary)
	Metropolitan Freight Survey (ordinary) (Tokyo, Kansai, Chukyo)
	Person Trip Survey (ordinary) (Tokyo, Kansai, Chukyo)
Road	National Road and Street Traffic Situation Survey (Road Traffic Census) (ordinary)
	Survey of Roads (administrative)
Car transport	Survey on Motor Vehicle Transport (fundamental)
	Survey on Motor Vehicle Fuel Consumption (ordinary)
	Mainline Bus Passenger Flow Survey (ordinary)
	Other Administrative Statistics, Statistical Data, etc. (28 statistics in total)
Railway transport	Survey on Current Rolling Stock Production (fundamental)
	Survey on Railway Transport (ordinary)
	Other Administrative Statistics, Statistical Data, etc. (17 statistics in total)
Port	Port Survey (fundamental)
	Other Administrative Statistics, Statistical Data, etc. (22 statistics in total)
Shipping, ships, and sailors	Statistical Survey on Coastwise Vessel Transport (fundamental)
	Survey on Shipbuilding and Engineering (fundamental)
	Survey on Seaman's Labor (fundamental)
	Mainline Ferry and Passenger Ship Passenger Flow Survey (ordinary)
	Comprehensive Survey on the Supply and Demand of Sailors (ordinary)
	Other Administrative Statistics, Statistical Data, etc. (39 statistics in total)
Aviation	Statistical Survey on Air Transport (ordinary)
	Airplane Passenger Survey (ordinary)
	International Airplane Passenger Survey (ordinary)
	Air Cargo Flow Survey (ordinary)
	International Air Cargo Survey (ordinary)
	Other Administrative Statistics, Statistical Data, etc. (5 statistics in total)

Source: The table was created by the author from the MLITT's Japanese website (statistics). http://www.mlit.go.jp/statistics/details/index.html
Note: "Fundamental" denotes Fundamental Statistical Surveys, "Ordinary" denotes ordinary statistical surveys, "administrative" denotes administrative statistics, and "processed" denotes processed statistics.

Statistics on the demand side rely on statistical surveys to collect data such as movement of people and goods, purposes of such movements, origin and destination, and means of transportation. With these statistics, researchers can analyze trends and predict future conditions. The Metropolitan Transportation Census and the Person Trip Survey, two of the surveys that fall into this category, are concerned with the movement of people. These surveys make it possible to obtain such information as purposes of trips and use of multiple transport means, which cannot be obtained from the supply side of transportation. Meanwhile, the Net Freight Flow Census in Japan provides information on the movement of cargoes from the owners of the cargoes. The survey approaches to businesses in various industries with large shipments of cargoes gather data on movement of cargoes in broad areas by investigating the types of goods, points of origin and destinations, and means of transportation for the domestic transportation of goods like raw materials, products, and waste.

As these statistical surveys have been conducted only in metropolitan areas or certain business sectors, there are areas or sectors where data have not been collected. The surveys also ask respondents to answer their detailed movements, making the response rate very low and increasing the effects of non-sampling errors. These statistical

surveys are classified as ordinary statistical surveys not mandatory for response because of the large burden on respondents and the impracticality of imposing an obligation.

In planning and designing the surveys given above, the statistical divisions in charge of the surveys normally establish an expert group consisting of users in academia and policy departments and invite suggestions for improvement. Such a process of interaction between producers and users of statistics helps enhance relevance of surveys in scientific researches and policy formulation. The same schemes are often used in other fields of transportation statistics.

3.2. Statistics on traffic accidents and safety

Of all the various traffic-related accidents, the National Police Agency (NPA) creates and publishes statistics on road traffic accidents. NPA also publishes statistics concerning traffic safety, including figures detailing traffic volume and traffic congestion on major roads. For non-road traffic accidents, the Japan Transport Safety Board of the MLITT creates and publishes statistics for aviation, railway, and ship accidents. Marine accidents fall into the jurisdiction of the MLITT's Japan Marine Accident Tribunal, which publishes statistics accordingly. These statistics are "administrative" by nature, products of the information that the relevant government offices obtain for their respective operations.

Detailed statistics on road accidents and related matters are created and published by the non-profit institution, the Institute for Traffic Accident Research and Data Analysis (ITARDA), which was established for conducting comprehensive analyses of traffic accidents to contribute to the planning of traffic safety measures, with the cooperation of NPA and the Prefectural Police. ITARDA publishes various research reports by conducting detailed analyses into the causes of traffic accidents, and also tabulates tailor-made statistical tables upon request.

3.3. Energy and environment

Statistics on energy consumption in the transportation sector are recorded under two heading items, passengers and cargoes, on the "Energy Balance Sheet," comprehensive energy statistics compiled by the Agency for Natural Resources and Energy Agency. Among environment-related information, the Ministry of the Environment records the carbon dioxide emissions of the entire transportation sector in its report "Greenhouse Gas Emissions Data of Japan." Estimated by combining with other basic statistics, these sets of data both fall into the "processed statistics" category.

To offer data regarding air pollution on roads, the MLITT and local governments also release hour-to-hour observation values of pollutants for major road locations on their websites [4, 5].

3.4. Statistics on the movement of people

There are four major factors causing traffic of people: (i) tourism, (ii) immigration, (iii) commuting flow, and (iv) residential migration.

Statistics on tourism come primarily from the Japan Tourism Agency (JTA) of the MLITT. The JTA collects statistical data from the following surveys, all of which are ordinary statistical surveys:

National Tourism Survey
Accommodation Survey
Consumption Trend Survey for Foreigners Visiting Japan
Regional Tourism Economic Survey

The Japan National Tourism Organization also publishes some statistics on tourism.

For statistics on immigration, the Immigration Bureau of the Ministry of Justice publishes the number of inbound and outbound Japanese travelers (by airport and seaport) and that of inbound and outbound non-Japanese travelers (by nationality, resident status, and other criteria) in the monthly "Statistical Survey on Legal Migrants," a set of administrative statistics obtained as part of the administration of immigration control. In addition, the Japan National Tourism Organization processes statistics of the "Statistical Survey on Legal Migrants" and publishes its calculations of the numbers of foreign tourists visiting Japan under a concept more appropriate for analyzing tourism.

Commuting flow, which consists mostly of movement that occurs on a daily or weekly basis between homes and offices or schools, is a major factor causing regional traffic. Statistics on the commuting flow on specific days is available in the Metropolitan Transportation Census and the Person Trip Survey, and they offer some but limited possibilities for analysis.

In contrast, the "Population Census" is a complete-count survey taken every five years. It includes statistics on regular commuting flow of workers and students within and between municipalities by origin and destination across municipalities. The Population Census is a "Fundamental Statistical Survey" because of its nature as one of the central pillars of the statistical system, not only providing a basis for formulation of transportation policies but also in supporting a variety of other public policies and offering benchmark data for other statistics.

Residential migration is also one of the causes of traffic, although not nearly as frequent a contributor as commuting flow. The Statistics Bureau of the MIC publishes on the monthly and yearly basis the "Report on Internal Migration Derived from the Basic Resident Registrations," a set of administrative statistics that uses information on municipalities' Basic Resident Registration. The Population Census also investigates locations of previous residences five years prior to the census reference date, making it possible to gather information on population movement over those five years by origin and destination.

4. Challenges and the future direction of transportation statistics

4.1. Current conditions and challenges

For better understanding the actual conditions of transportation and traffic, statistics produced from various methods, including survey statistics, administrative statistics, and processed statistics are needed. Because of the public nature of the transportation and traffic sector and the sizable role of the government in the field, Japanese transportation statistics include a plethora of administrative statistics based on administrative documents and they are supplemented by various kinds of survey statistics. Thus, a wide variety of detailed statistics are made available for researches, and the current methods for producing statistics are quite reasonable in view of the characteristics of the transportation and traffic sector. Availability of transportation statistics is quite similar between Japan and many Western countries, and the methods for producing the statistics are quite common between Japan and the Western countries.

With the needs of the policymakers and researchers who use the statistics becoming increasingly sophisticated and diverse, however, the statistical offices are always required to provide statistical data that is more detailed and accurate, and easier to use to help elucidate the latest social and economic issues. Statistical offices have to produce and disseminate statistics as timely and efficiently as possible, and those specialized in transportation statistics are no exception.

Considering the above background, two approaches for developing Japanese transportation statistics will be needed; one is to enable more effective use of the existing statistics and the other is to develop new statistical sources for the transportation statistics.

One measure in the former approach is to make an environment where microdata are accessible to researchers. Formerly, statistical offices produced and disseminated statistics in the form of tables that would be commonly needed by many users. However, this makes it difficult for users to analyze data other than the published statistical tables. In other cases, users may need to use microdata directly for

multivariate analyses. To meet such research needs, it is important to develop an environment where researchers can access microdata.

For the latter approach, one possible way to create new sources of information for the transportation sector would involve creating and analyzing statistics based not only on administrative data and survey data but also on "big data" that come from the Internet, mobile phones, various sensors, and other sources. The development of ICT has made it possible to measure and collect the data on positions and movement of vehicles and individuals in higher frequencies and granularities without making substantial investments of time or money. Using this kind of information in an effective manner will make it possible to respond more suitably to the recent issues of transportation and traffic.

The following sections discuss these two issues in more detail.

4.2. Using microdata

In this section, general status of use of microdata in Japan will be explained first, and then the issues concerning microdata for transportation statistics will be discussed.

Prior to the revision of the Statistics Act in 2007, the emphasis of the law was on the protection of confidentiality of the survey data at the cost of data usability. At the time of the enactment of the first Statistics Act of 1947, electronic computers were still at a primitive stage of development, and it was almost unthinkable that many statistical users would be able to have their own computers for their analyses. In those days, therefore, the aggregated statistical tables were taken for granted as the only form of disseminating official statistics. Therefore, lawmakers did not consider making a provision to allow access to microdata; users could get access to individual data, or microdata, only in exceptional cases.

As computers proliferate, and many researchers came to use them, demands for more detailed tabulated results increased. Researchers have better potentials to make more in-depth analyses if they can perform, for example, multivariate analysis directly on the individual data. To accommodate to such new demands for researches, the Statistics Act was revised in 2007 to allow access to microdata for research purposes.

The new Statistics Act allows researchers who want to use Microdata to choose from the following three options: (a) receiving a customized tabulation service for fees, (b) using the "Anonymized Microdata File" created from the individual data of the statistical survey, or (c) using the individual data for researches with significant public interest, such as joint research with a government agency.

Those who select option (a) cannot access the individual data directly, but they can request the statistical offices to produce aggregated tables by specifying the format. This option is simple and convenient, when the data table in need is not available in published statistics.

Under option (b), the statistical office creates an anonymized file, eliminating all identification information (name, address, and household number, etc.) from the individual data files, and applies an additional anonymizing procedure. User can obtain such files by submitting a simple application form. The anonymization process makes it practically impossible to identify individual respondents from the file, ensuring confidentiality of respondents' identity.

Under option (c), access to identification information is usually restricted, but researchers would still be able to perform analyses using most of the survey information prior to the anonymizing procedure. As option (c) makes it easier to identify the respondents than option (b), however, it requires strict measures for protecting confidentiality. For this reason, to be eligible for option (c), a research project must thus be a joint research initiative with a government agency and/or constitute a project of a public nature. In this case, users must obtain permission from the statistical office that conducted the statistical survey.

Currently, eight ministries are providing option (a) services for 26 surveys. For option (b), the Statistics Bureau of Statistics of MIC and the Ministry of Health, Labor and Welfare are providing anonymized data files for 7 surveys. The anonymization process for this option is technically complex and challenging, which limits the number of statistical surveys for which this option is applicable. This option is available only for household surveys, and it is generally difficult to apply to surveys of corporations and establishments. This is because there are usually some very large companies, and they can be identified from some variables, even if all the identification information were deleted. This is why the number of statistical surveys using option (b) is relatively small.

Under the framework of the Statistics Act, the number of cases where microdata was used in transportation statistics, most of which are under the jurisdiction of the MLITT, is relatively small compared with statistics in other fields. One reason for this is that options (a) and (b) are not yet provided for transportation statistics. As for option (c), there were 14 cases applied to transportation statistics in FY 2013. Compared to the total of 244 cases of option (c) for all the official statistics, the number for transportation statistics constitutes only a small proportion [6]. One way to interpret this low figure could be that the users were satisfied with the option (a) aggregated statistical tables. However, a more plausible explanation would be that there would be a strong potential need for the use of microdata that had not been satisfied. In particular, the statistics on traffic flow, such as the Person Trip Survey, may present opportunities for more diverse analyses because they contain information on the origin–destination movement of trips that can be used in detailed analyses from a wide array of viewpoints.

Another reason for the low level of microdata use in the transportation statistics is that a large number of transportation statistics are administrative statistics. Unlike the use of microdata for survey statistics, there is no unified framework regulating the use of microdata for administrative statistics because administrative statistics are subject to the specific laws governing the administrative operations, not the Statistics Act. This factor also complicates the use of microdata in transportation statistics.

Although the revisions to the Statistics Act have opened the doors for the use of microdata, there is still much room for improvement for using microdata. In the near future, the statistical offices creating survey statistics will have to review their overall track records of providing microdata files and make more efforts to enhance convenience and usability. In addition, for survey statistics for which microdata are not yet provided, the relevant statistical offices will have to start creating anonymized microdata learning from the practices of the precedents of other surveys.

4.3. Using big data

The high cost of statistical surveys and the difficulty of ensuring statistical accuracy are two main issues of current transportation statistics. For example, the 5th Tokyo Metropolitan Area Person Trip Survey (2008) employed a random sample of 1.4 million households selected from the 16 million households in the Tokyo metropolitan area [7]. Because it employs a mail-out and mail-back method, the cost of conducting the survey is lower than the interview method. But still, the cost of the survey is substantially large. It is difficult to ensure high accuracy because the response rate is quite low. In the same survey, the effective response rate was 24%. For example, the response rate for the fifth Person Trip Survey for Tokyo Metropolitan Area (2008) was 24% [7]. This is attributable to the large response burden resulting from detailed survey questions. Although the office in charge of creating the statistics makes adjustment for non-response in the estimation of the results, the effect of non-sampling error remains to be significantly large.

One effective method for solving this type of problem is to use "big data" as a new source of information. While there is no universally agreed definition for "big data," it is generally considered as collections

of data compiled electronically from the Internet, GPS, and various sensors tracking movement of people and goods. For example, public transportation operators now collect fees via magnetic tickets or prepaid IC cards and can use the big data of the movement of passengers. If the system for collecting passenger fares were to rely on traditional paper tickets, operators would have much work to do to compile statistics on passenger flows. However, the automatic fare collection system makes it possible to obtain detailed information on the flow of passengers by tabulating and analyzing the electronic records without additional data collection.

In addition, there are also other information sources for the general movement of people and vehicles. For example, mobile phone systems collect location information of the phone users, and automobile navigation systems collect location of automobiles continuously. If this type of data can be used for statistical purposes, detailed statistics on transportation can be produced in a specific field more accurately and efficiently.

Thus, big data have a high potential to improve compilation of transport statistics and researches on transportation problems dramatically, not only in Japan but also in other countries. Developed countries generally have good advantage over developing countries in using big data for analyzing traffic and transportation problems because in such countries, transportation systems are quite often supported by advanced information and communication technologies (ICT), and there are many potential data sources to be used for various analyses. Developing countries, however, are not always severely disadvantaged in using big data for traffic and transportation analyses because some developing countries are implementing new ICT-based systems, such as traffic control systems, and such countries will have opportunities to access to the big data collected by such systems. Even in countries where ICT-based traffic and transportation systems are not yet popular, it is still possible to use data on location and movement of people collected in the mobile phone systems as big data. Big data, therefore, provide new approaches for expanding data sources in both developed and developing countries.

Despite the high potential of big data on traffic and transportation, there are certain problems that impede the practical use of big data. To resolve such problems, it will be necessary for policy makers and researchers to share experiences among sectors and across countries.

One issue is that big data are mostly proprietary information belonging to particular companies. Unlike official statistics, corporations have no obligation to share the proprietary information with the public. For example, leading e-commerce companies collect and analyze the data on their own customers browsing and purchasing behaviors as proprietary secret because such data sets are the source of competitiveness for such companies. For mobile phone operators, data of their customers cannot be shared with a third party for the same reason, and also for protecting the privacy information of their customers. Therefore, it is difficult for anyone outside such companies to access and utilize the information.

Another problem of the big data is that a particular set of big data does not necessarily represent the entire statistical population. For example, the location information from mobile phones is highly accurate in general, but the information from one mobile phone company represents only a part of the total population. One person may subscribe to two or more mobile numbers, or a mobile phone user may sometimes switch off, and location information would not be available. For these and other reasons, the total population aggregated from the mobile phone information does not agree with that of the population figure obtained from official statistics. To resolve this kind of problems, it is necessary to perform detailed analyses about the relative size and characteristics of the mobile phone subscribers in relation to the total population. Some Japanese mobile phone companies are currently researching the possibilities of estimating the movements of people based on mobile phone location information. In order to apply big data to transportation statistics, researchers will need to develop an estimation method and create auxiliary data for estimating populations from big data.

There is also the key issue of protecting personal privacy. For example, examining the daily location information from an individual's mobile phone reveals the person's movement on the day in question, and from such information, the person's behavior can be inferred. Should a company provide its big data to a third party, the company will have to face with the significant risk of exposing individuals' information. The use of big data for research will require information management rules for strictly protecting privacy.

In recent years, an increasing number of people disapprove the idea of companies providing their personal information to a third party against their wishes, and news media often report such incidences. One well known case in Japan is the one of East Japan Railway Company (JR East) in June 2013. The company provided passenger data to an information processing company to develop better ways of utilizing the traffic and transaction data of the prepaid IC card. Many JR East passengers use prepaid IC cards called SUICA that can be used for paying the train fares at the gates and purchasing goods in many stores having the IC card readers.

JR East provided the data to a data processing company under an agreement to keep the data confidential and data protective measures. The individual IDs in the file were transformed into a set of completely different codes that cannot be converted back to the original IDs. Also, the date of birth was limited to the month and the year. With these protective measures, from the technically point of view, the transformed file was virtually free from the risk of identifying individual card holders. When the news was reported in the media, however, harsh criticisms arose from the users of SUICA against the provision of the personal data to the third party.

As this case suggests, utilization of big data has still a very short history, and the rules for handling such data are not yet well developed and socially agreed. This case made the handling of corporate information about customers a highly sensitive social issue.

In light of these incidents, the Japanese government prepared a bill to amend the Act on the Protection of Personal Information in March 2015 and plans to submit it to the Diet in the near future. It aims to give a new framework for treatment of corporate data. It is based on the idea that the effective utilization of such data will generate significant social value benefiting the customers themselves as well [8,9]. Under this scheme, users will be able to make more effective use of big data in the future.

Aside from the corporate data, it should be noted that not all forms of big data face the same difficulties as the aforementioned privacy issues and the population estimates. In collecting data on road traffic conditions, for example, users can measure various data by mounting drive recorders to vehicles such as buses and taxis, which are often called "car probe data." If researchers can access data on the speed and congestion from such a data source, it would make useful information. In this kind of cases, privacy concern does not become a problem for the vehicles owned by the public body or in the cases where prior consent of the vehicle owner is obtained. Studies have already been under way for collecting and analyzing this type of data toward practical use.

While there are some cases where big data began to be used for researches on transportation and traffic problems, there is still much to be done to utilize big data for research and statistical purposes. Although it seems to be quite difficult to replace current set of transportation statistics with big data completely, big data will have significant roles in complementing the shortcomings of official statistics or adding valuable information not available from conventional data sources. Big data are expected to contribute to various aspects of statistics and researches, such as reduction of the cost and the response burdens of data collection and improvement of timeliness and granularity of the data.

Technological innovations are changing the cost and quality of data collection significantly, and the research communities should take

advantage of this opportunity for enriching the data sources of transportation statistics. To that end, the government, the academia, and the private sector should cooperate in exploiting the wide variety of information sources so that the transportation statistics will meet the needs of society better.

5. Conclusion

Although the existing transportation statistics are fairly developed in Japan, there is still much room for improvement. In particular, efforts will be needed to extend the use of microdata for supporting scientific and policy researches and to use big data as new data sources for statistics and researches. Such measures will benefit all the parties involved, i.e., the government, the academia, the private sector, and the public. It is necessary to construct a system that will make effective use of information for the betterment of society by utilizing new technologies while protecting privacy information. To complement the shortcomings of statistics compiled from surveys, big data have a high potential as new sources of information. It is still necessary for most countries to accumulate experiences of using big data for analyses of traffic and transportation problems. International exchange of such experiences will help develop use of big data as a new type of data sources.

References

[1] Ministry of Internal Affairs and Communications, Director-General for Policy Planning (Statistical Standards), Statistics Act, http://www.soumu.go.jp/english/dgpp_ss/seido/1-1n.htm.
[2] Statistics Bureau, MIC, Portal Site of Official Statistic of Japan (English version), http://www.e-stat.go.jp/SG1/estat/eStatTopPortalE.do.
[3] United Nations Statistics Division, Fundamental Principles of Official Statistics, http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.
[4] Ministry of the Environment, Atmospheric Environmental Regional Observation System: AEROS, http://soramame.taiki.go.jp/Index.php.
[5] Kanto Regional Development Bureau, MLITT, http://www.ktr.mlit.go.jp/road/shihon/road_shihon00000034.html.
[6] Ministry of Internal Affairs and Communications, Statistics Act Enforcement Status Report 2013 (in Japanese), http://www.soumu.go.jp/menu_news/s-news/01toukatsu01_02000034.html.
[7] Tokyo Metropolitan Area Transportation Planning Council, Person Trip Survey", http://www.tokyo-pt.jp/person/.
[8] Masashi Kubota: "Utilization of personal data and the amendment of Personal Information Protection Act"(In Japanese), Legislation and Research 2015.1 No.360, House of Councillors Secretariat Planning and Coordination Office, Japan, 2015.
[9] Cabinet Secretariat, Bill to amend the part of the Act on the Protection of Personal Information and the Act on the Use of Numbers to Identify a Specific Individual in the Administrative Procedure, Bill submitted to the 189th Ordinary Diet, http://www.cas.go.jp/jp/houan/189.html (accessed March 15, 2015).